Homework 1: Foundation Concepts (I)

COMS W4111: Introduction to Databases  
Sections 002  
Spring 2025  
  
(v 0.75; 2025-JAN-26)

**Note: This document contains part “a” for homework 1. You may get started. Completing the full homework will require material from lecture 2. We will release the complete HW after lecture 2.**

# Submission and Overview

## Submission

Total points for homework assignments and exams determine final grade. The final point total is between 0 and 100. HW 1 is worth 5 points.

**Due date: 2025-Feb-09, 11:59 PM EDT on GradeScope.**

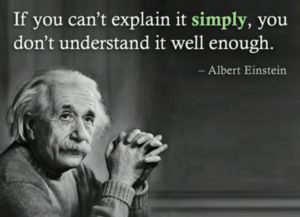
Please note:

* You submit on GradeScope.
* There is a [post/mega-thread](https://edstem.org/us/courses/73023/discussion/6015704) on Ed Discussions that contains submission instructions. We will use this thread to answer questions and clarify the assignment.
* You may start the homework when you want. Some of the questions will require material from lecture 2.
* The scope of material for this homework is:
  + The material in lecture 1.
  + The material in lecture 2.
  + The slides associated with the recommended textbook for
    - Chapter 1.
    - Chapter 2.
    - Chapter 3, slides 3.1 to 3.36.
    - Chater 6, slides 6.1 to 6.24.

You submit two artifacts when you submit the homework assignment.

1. A PDF of this document with answers to the questions.
2. A PDF export of the Jupyter Notebook for this homework. The notebook is in the same folder as this document and is S25-W4111-HW1.ipynb. The post of Ed will explain how to create a PDF of a notebook.

## Brevity



Keep your answers focused, brief and succinct. The answers to written questions only require 3 or 4 sentences/bullet points. If you ramble and bloviate hoping to get something correct, we will deduct more points.

## General Knowledge Questions

1. Is data in a spreadsheet unstructured, semi-structured or structured. Briefly explain your answer.
2. Professor Ferguson can export SSOL data for his classes to a spreadsheet. Despite the data being spreadsheet-like, Columbia uses an application and database to manage the data. List 4 functions/capabilities of a database management system that simply sharing spreadsheets lacks.
3. What are the three levels of abstraction for data that a DBMS provides? What are two disadvantages of having a user or developer directly use the lowest level, direct access to data?
4. Is a full stack web application a two-tier or three-tier application? Is Jupyter notebook a two-tier or three-tier application? What say a two-tier application product in lecture 1. What was the product?
5. What are the four types/categories of database user? Which type of user is most likely to use DDL?
6. Consider the entity set with relationships below. Assume that the bold attribute/column is the primary key. Write the relational model schema definitions for representing the entity sets *Faculty* and *Section.*A diagram of a graph

   AI-generated content may be incorrect.
7. Using the approach for documenting relationship sets in Lecture 1’s slides, write down the relationship set *Teaches* for the diagram in question 6.
8. For the diagram in question 6, draw the *conceptual model* Crow’s Foot Diagram using [Lucidchart](https://www.lucidchart.com/pages/landing).
9. What is the result of the relational algebra expression Π course\_no, course\_title (Section)?
10. What is the SQL equivalent to the relational algebra expression π name, dept\_name (σ dept\_name='Comp. Sci.' (instructor)).